DOCUMENT RESUME

ED 066 445 TM 001 775

AUTHOR	Kennedy, E. G.; And Others
TITLE	Position Paper on Standardized Testing.
INSTITUTION	Kansas Advisory Council for Guidance and Pupil Personnel Services, Topeka.
SPONS AGENCY	Kansas State Dept. of Education, Topeka.
PUB DATE	2 Feb 72
NOTE	15p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Evaluation Techniques; *Measurement Instruments; *Standardized Tests; Statewide Planning; *Student Testing; *Test Construction
IDENTIFIERS	*Kansas

ABSTRACT

Recommendations of an Ad Hoc Advisory Committee relating to standardized testing in a state educational system are presented. The paper first discusses the concepts of measurement, evaluation, and standardized testing. Then follows discussions of Test Development, Qualifications of Test Users, General Use of Tests, The Use of Standardized Tests for Individual Assessment, and The Use of Standardized Tests for Program Assessment. It is recommended that: (1) each local educational agency establish systematic procedures for planning, implementing, and evaluating the testing programs within the LEA; (2) a permanent committee be established with the responsibility of examining state-wide issues concerning testing and/or individual and program assessment. (DB)

ED 066445

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

"POSITION PAPER ON STANDARDIZED TESTING"

Prepared
by

STATE ADVISORY COUNCIL
for
GUIDANCE AND PUPIL PERSONNEL SERVICES

February 2, 1972

E. G. Kennedy, Chairman
Brooke B. Collison, Chairman,
Ad Hoc Committee

The following paper is the result of a charge from the Executive Council of the State Department of Education to the State Advisory Council for Guidance and Pupil Personnel Services on January 1, 1971. The charge emanated from concerns about school testing which have been reflected in popular press and professional literature of recent years regarding the appropriateness of tests and their appropriate uses. This work was completed on February 2, 1972.

The attached recommendations are considered to be a separate part of this position paper and are directed to the Executive Council of the State Department of Education for their consideration only if and when the philosophy of this paper becomes a recommended part of the educational process.

TM 001 775

4/15/71

GUIDANCE AND PUPIL PERSONNEL SERVICES ADVISORY COUNCIL

AD HOC TEST ADVISORY COMMITTEE

Dr. Donald R. Bloss
Counselor Education
Fort Hays Kansas State College
Hays, Kansas 67601
Ph. (913) 628-4000

Dr. Brooke B. Collison, (Chairman)
Counselor Education
Wichita State University
Wichita, Kansas 67208
Ph. (316) 685-9161 Ext. 441

Mr. Larry Gaston
Director of Research
Topeka Public Schools
415 West 8th Street
Topeka, Kansas 66603
Ph. (913) 357-0351

Dr. E. G. Kennedy
Psychology & Counselor Education
Kansas State College of Pittsburg
Pittsburg, Kansas 66764
Ph. (316) 231-7000 Ext. 344

Mr. Donald M. Moritz
Library Building
Sixth and Ann Streets
Kansas City, Kansas 66101
Ph. (913) 621-3073

Mrs. Margaret Newbanks
Director of Instruction
Parsons District Schools
909 South Central, Box 377
Parsons, Kansas 67357
Ph. (316) 421-5950

Dr. Dale P. Scannell, Dean
School of Education
University of Kansas
Lawrence, Kansas 66044
Ph. (913) UN4-3726

Dr. Ralph Walker
Director of Research
Wichita Public Schools
Curriculum Building
640 North Emporia
Wichita, Kansas 67214
Ph. (316) 268-7882

Mr. Richard Warren, Counselor
Leavenworth High School
Tenth and Halderman
Leavenworth, Kansas 66049
Ph. (913) 682-8653

POSITION PAPER ON STANDARDIZED TESTING

Certain premises undergird the practice of standardized testing. Chief among them is the premise that testing is a part of the measurement and evaluation process within education. Furthermore, it is the responsibility of all interested parties in education to develop the most valid and effective measurement and evaluation scheme possible within reasonable limits of time, economics, and personal rights. Indeed, to do less may result in subsequent educational decisions being similarly less effective and/or less valid.

The concepts of measurement, evaluation, and standardized testing need to be defined more explicitly at this point so that they may be seen in their proper perspective within the total educational system.

Measurement is a process of assigning numeric or quantitative descriptions to particular traits, characteristics, or behaviors which have been observed. Neither a numeric description or any part of the measurement process has value in and of itself.

Evaluation, therefore, is the process of assigning value to the measurements collected. Measurement can exist by itself, but evaluation must be preceded by measurement. Evaluation is only as effective as the measurement on which it is based and is no more valid than are the traits being measured for the person, persons, or tasks under consideration. Similarly, measurement and evaluation in education are appropriate, only to the degree that measurement procedures are related to evaluation objectives.

Standardized tests, according to the Ad Hoc Test Advisory Committee, are published instruments designed to measure certain traits, characteristics, skills, attainments, or potentialities which are a part of the individual at some point in the education process. Furthermore, the use of the word "standardized" should be reserved for those instruments which have met acceptable criteria of development and use. A premise on which the Ad Hoc Test Advisory Committee has functioned

is that the capabilities and limitations of tests selected must be known by the user before any use can be legitimately made of any standardized test.

Contained within this premise are two assumptions: (1) that test companies have researched their own instruments carefully and have made that research available to users; and (2) that users have acquired adequate skills requisite to the various tasks required of them in the use of standardized tests. These assumptions are consistent with discussions within the Ad Hoc Test Advisory Committee which repeatedly concurred with the statements that test abuses stem from inappropriate instruments or improper use of instruments, either of which could be avoided if additional test information were available or if users were more sophisticated psychometrically.

The test publisher's responsibility for adequate research, development, distribution, and control of standardized tests has been adequately defined in an American Psychological Association publication, Standards for Educational and Psychological Tests and Manuals. The Standards have been available since 1966 and have been used by test reviewers and others whose responsibility it is to comment on tests which are commercially available. Unfortunately, the Standards, like many guidelines, may be ignored with impunity; and, although prepared by a most highly-qualified committee representing three professional associations concerned with testing, the Standards are still judgments and recommendations with which persons are free to disagree. There does not exist an effective procedure for redressing grievances by those tested of psychometric instruments outside of personal appeal to the courts.

The committee position has been that the system of public education should encompass all youth. Within that system, standardized testing is an appropriate tool which can be used to assist all individuals. This ideal can become a reality only when the education system operates from an adequate information base and can adjust school experiences in view of accurate information. Descriptions of

the individual, the system in which he operates, the society in which that system functions, and the interrelated objectives of each are essential parts of that information base. It would seem superfluous to state that for such a system to work, as much information as possible must be collected about the individual--where he is, where he is going, whether he is making progress, and, if so, at what rate of speed. The testing specialists must know as much as possible about the group to which the individual belongs. They must be able to assess the part that the educational system plays in either the individual's or the group's movement toward objectives. They must know what the society is like and how it is changing so that all persons can be contributing members of that society.

Effective attainment of the objectives just stated would be facilitated by (1) an adequate program of individual assessment; (2) an adequate program of assessment of the group within which the individual is a member; and (3) an adequate assessment of educational programs. Conventionally, assessment is thought of as being in the areas of ability, achievement, and interest. The essential reasons for the emergence of these three areas are that they represent common concern for educators, they represent common correlates of school achievement, and historically they have been the common areas of test development which have met with success. This is not to say that additional areas of concern are unknown nor that they are unimportant; in fact, new assessment techniques in the non-cognitive domain may be needed more. The simple point is that assessment devices in the cognitive domain have been easier to construct and have been the ones that "sold."

Even were test publishers to market only instruments which had met the most rigorous developmental standards, errors of use would still be possible. Indeed, the burden of credit or guilt for either appropriate application or abuse through inappropriate application must rest with the person responsible for using tests, regardless of the situation. The number of actual or potential

abuses are too numerous to list and nearly impossible to itemize. The generalization remains that the best safeguard for appropriate test use is a well-trained user. The committee's position is that there is no single level of training appropriate for all persons: the several levels of use (classroom teacher, school counselor, clinical psychologist, or statistical researcher) demand uniquely specific preparation, the need for which must be determined in part by the professional groups represented.

Additional or different training on the part of the professional person who uses tests will not guarantee appropriate use, nor will it eliminate a more common misconception about standardized testing held by the lay public--that a single test administration can be analytic and diagnostic of isolated factors which are confoundingly imbedded in multiple cause and effect. School achievement is a product of multiple causation; the single score on a test taken at any given time can only be a sample of a specific behavior at that particular time. The score is a function of the interaction of many variables: the pupil's ability, curricular materials, quality of instruction, environmental influences, previous development, previous experience, conditions of the testing situation, condition of health, and/or many other possible factors. Consequently, the single score cannot be representative of any single causative element, but it is representative of the sum and interaction of all. A score on a single instrument, therefore, cannot be used as a criterion for an imbedded causative element without carefully controlled conditions having been designed to isolate the cause.

It must also be remembered that school testing programs are concerned with factors which are emotionally loaded: the ability, achievement, and interest of children. Paramount among the considerations in any testing situation must be the rights of the individual being tested. Respect for the individual and his personal privacy, as well as respect for the beneficent use of test data obtained from an individual, must undergird all psychometric practice. Schools

should discuss and carefully plan for the respectful collection, storage, and dissemination of pupil data. To do otherwise is to invite test abuse and public criticism. The criterion for respectful use should be that test use should benefit, not harm, an individual. In this same vein, much more needs to be done by professionals in testing to communicate appropriate and accurate information to the lay public about test purposes and test capabilities.

TEST DEVELOPMENT. Although the proper use of a test or its misuse ultimately is the responsibility of the user, the test publisher (both commercial and other) must bear responsibility for providing a test manual that contains a full description of the rationale for the test, stated so that there is no doubt in the user's mind to what subjects and to what group the test is related. The manual should contain a full and relevant description of all of the types of validity on which information is available for the test. The manual should contain an appropriate description of the test, re-test, parallel form, and/or internally consistent reliabilities, including all reliability coefficients. It should contain a complete demographic description of norm samples and each population sample used in standardization. Separate norms by racial, ethnic and economic sub-groups should be encouraged. When terms such as National or Population norms are used, the samples should be large, representative of the population, and fully described. Dates of standardization and norming trials should be clearly indicated.

The test publisher has the responsibility of listing the limitations of the test as seen by the publisher and/or the authors. Inadequacies and inappropriateness for certain situations and for certain groups should be cited. In addition, the test developer has the responsibility for making the instrument as usable by the consumer as possible. This includes some concern for test length, time limits, scoring procedures, formats of answer media, and methods of reporting and recording. Careful decisions regarding these points not only aid in administration but help to eliminate student errors in taking the test,

as well as administrator errors in scoring, reporting, and recording.

QUALIFICATIONS OF TEST USERS. Those who use tests or test results should possess certain expertise which qualifies them for the level of use or interpretation at which they are going to use the test. They should be able to assess the strength and limitations of the instrument. They should understand the consequences of improper administrations, scoring, and interpretation of tests and test results. Test users should have had courses or in-service training in educational measurement which give them a basic understanding of the statistical characteristics of the test at the level to which those statistics are necessary for administration and interpretation. Those who do research with tests should have a thorough knowledge of statistics and statistical manipulation of test results. In addition, the users of tests would have an understanding of the confidentiality of the scores and the potential danger to the individual of improper use of test results. Also, the administrator, scorer, interpreter, and researcher of tests and test results should operate from a basic position of ethical behavior as outlined in professional ethical codes.

GENERAL USE OF TESTS. It is recognized that there are many different kinds of tests. A test with a definable purpose has merit, provided its construction and use are valid and ethical. Within broad limits and with certain considerations, tests may be used to identify levels of ability, aptitude, achievement, and interest. Tests that identify these factors are commonly used types of tests that are normally administered in group situations. Though there are other tests that are administered on a one-to-one basis, they will not be considered specifically according to type, use, and construction in this paper.

Tests may legitimately be used to assist in the evaluation of programs and traits of individuals. Further, tests may be used within broad limits for certain kinds of predictions for both groups and individuals. The right test in the right setting may be used as a tool for research. Tests have value for placement,

employment, and admission; however, these uses of tests would be made only when other verifying evidence is available. Tests probably become discriminatory when used as the only criterion for placement, employment, or admission of members of minority, economically deprived, and culturally different groups. Though much can be said on this specific subject concerning past misuse of tests, the fact remains that the appropriate test used ethically may constitute a legitimate aid to placement, employment, or admission rather than an instrument to prohibit them.

THE USE OF STANDARDIZED TESTS FOR INDIVIDUAL ASSESSMENT. When all aspects of tests, their use and misuse, are applied to individuals, the problems become multi-faceted for it is extremely difficult to discuss all aspects of test development, use, and interpretation without becoming redundant and narrowly limited with respect to any area. Many of the areas for discussion with individuals pertain also to program assessment and evaluation while others become unique to the test relationship with the individual. This paper will refrain from mentioning the former even though it is recognized that they exist.

It is impossible to discuss in this document all of the many kinds of tests--their development, use, and interpretation. Though some general guidelines may be developed, it must be stated or implied that an appropriate instrument must be used by an appropriately trained person for an appropriate purpose. Less than this is unacceptable.

THE USE OF STANDARDIZED TESTS FOR PROGRAM ASSESSMENT. Program assessment is the process of determining whether programs achieve desired results--the process of gauging program effectiveness. It is the determination of the outcomes of education interpreted in light of program objectives and of community and pupil characteristics. It includes the evaluation of courses, curricula, and programs. The goal of program assessment is to provide reliable and meaningful information over a period of time with a view toward improved educational

decision-making. The process includes the identification of areas of concern, the selection and collection of appropriate kinds of information, and the analysis and interpretation of data that are timely and relevant.

The process of determining whether educational programs acnieve desired results presupposes measurement and evaluation. The Ad Hoc Test Advisory Committee has taken the position that standardized testing is a valid part of the process. This position is based upon four fundamental assumptions which follow: The use of standardized tests for program assessment assumes (1) that agreed-upon goals and objectives exist which can be translated into measurable enties; (2) that standardized tests either exist or can be developed which are capable of providing measures of the attainment of educational objectives; (3) that objective information is needed for decisions related to financial resource allocation, program modification, and the like; and (4) that information yielded will be used in affecting changes in areas needed.

Most authorities in educational measurement agree that information yielded by standardized testing is needed to improve instructional programs and to identify needs of student populations at the local school, school district, and state levels. The Ad Hoc Test Advisory Committee feels that there is a need for knowledge of student population characteristics such as ability, achievement, and interests. Standardized testing can provide needed descriptions by individual schools, groups of schools, school systems, geographic areas, and community types. Changes over given periods of time can be studied, and comparisons with existing norms can be made when appropriate.

Assessment will assist in evaluating the effectiveness of innovative instructional programs at strategic points in time. Federal program evaluation and determination of needs are based in part upon standardized testing. Such evaluation and needs assessment increase the demand for reliable and valid testing. Test scores provide information which can be used in cost-benefit analyses where

such scores represent products of a school or a school system. Information yielded by tests can result in (1) increased understanding of the outcomes and deficiencies of the schools; (2) better planning and direction at all levels; and (3) more and better assistance where needed. Potentially, improved legislation to meet educational needs can result from wise consideration of test results along with other relevant data.

There is much criticism of tests and testing with regard to their use and application both with individuals and groups. This criticism is leveled from within and without the education profession. The bulk of the criticism seems to relate to the use of tests with particular groups of individuals, e.g., minority groups, those with backgrounds of poverty, and others who are educationally disadvantaged. The criticism relates to two primary issues--whether tests measure what they purport to measure and whether tests are used the way they should be used. Another important issue, the consequences of testing, is akin to the latter. Certainly it seems possible that the same test might measure different characteristics in one group from what it measures in another group. This being the case, care must be taken in making and using interpretations of test results. Such interpretations must be made by competent persons. Decisions concerning test use must take into account not only the psychometric properties of instruments in question but also the specific purposes for the testing and the possible consequences, including side effects. Side effects may be either positive or negative. One example of a negative side effect of testing is the instance where negative feelings toward learning are reinforced by the testing process.

More and more reliance is being placed upon the use of standardized tests for program assessment in spite of the growing controversy about testing and the increasing number of questions raised about the validity of the tests used and the effects upon those who take them. The use of tests for this purpose

is increasing at the national, state, and local levels. The committee subscribes to such use, provided appropriate guidelines, such as the following are observed.

Testing for program assessment should be done only in amounts needed for determining the value of the program. A criticism which is sometimes valid relates to the amount of testing done in schools. Excessive testing should be avoided. Sampling should be utilized when possible and when the results of complete batteries on individuals are not needed for other purposes.

Tests should be utilized in accordance with their intended usage. Some tests may be appropriate for multiple uses. If test results are to be used for other than the original purposes, justification must be based upon scientific grounds as well as the potential social consequences. The limited number of characteristics capable of being measured by any one instrument must be recognized. Appropriate reliance should be placed on these while recognizing that there are important outcomes for which standardized instruments may not have been developed. Examples of such outcomes are certain feelings, attitudes, and appreciations.

Differences in the characteristics of groups being tested should be taken into account, especially when comparing the performance of various groups. As indicated above, educational program outcomes must be interpreted in view of differing financial resource levels and differences in community characteristics and students' backgrounds. Measures other than those related to student performance must be obtained and analyzed. Examples of such measures are those pertaining to socio-economic characteristics of the community and to conditions existing within schools.

Finally, test results should be used in making decisions regarding educational programs. Testing results should be used to help bring about change. There should be prior agreement regarding what will be done if test results reveal certain deficiencies. If tests being utilized in program assessment possess adequate psychometric properties, if they are used in legitimate ways, and if the

results are interpreted in light of relevant information, decisions will be more nearly valid and education will improve because of their use.

RECOMMENDATIONS

1. The philosophy of the Test Advisory Committee has been that the persons or agencies affected by the collection of test data should be involved in the planning of that test data collection. It is therefore recommended that each local educational agency establish systematic procedures for planning, implementing, and evaluating the testing programs withing the LEA. There is more than an implication here that the only legitimate use of test data is a "planned" use, with the plan normally made prior to collection of data.

2. It is recommended that a permanent committee be established by the State Board of Education which would be charged with the responsibility of examining state-wide issues concerning testing and/or individual and program assessment. This committee would be multidisciplinary in its make-up to reflect the affected groups involved in any state-wide program of assessment. It is felt that in addition to appropriate school personnel--teachers, administrators, counselors, board of education members--that parents and students could well be represented; however, a high level of technical competency should be present among members from the professional community.

   2.1 Tasks with which a state committee on testing might be concerned would be

      2.11 Examination of the need for individual or program assessment state-wide;

      2.12 Recommendations concerning procedures for individual and program assessment state-wide, if needed;

      2.13 Recommendations concerning dissemination of information resulting from state-wide assessment;

2.14 Control of access to information collected on a state-wide basis where requested by researchers or agencies whose requests were not cleared prior to data collection;

2.15 Survey of training standards which concern persons who use tests;

2.16 Preparation of guidelines for the collection and utilization of standardized test data with subgroups of the general population.